

# Chapter 7

## Conclusions

### 7.1 Conclusions

In this thesis, I describe the some of the first steps in the application of whole genome sequence data to study the exceptional diversity of East African cichlids. I generated an annotation of microRNA loci for five reference genomes, predicted which protein coding genes may be regulated by them, and explored sequence evolution both in the genes themselves and their targets. Then I took advantage of one of the reference genomes (the *M. zebra* genome for Lake Malawi) and aligned whole genome sequences from 239 individuals to it to obtain two detailed catalogues of variation: over 20 million genetic variants for Lake Malawi and almost 5 million variants to study incipient speciation of *Astatotilapia* cichlids in the isolated crater lake Massoko. Another 29 Lake Malawi and 4 crater lake region *Astatotilapia* individuals have been sequenced but their genomes have not yet been analysed. To further facilitate research on this fascinating system, I constructed whole genome alignments between the reference genome assemblies, assigned ancestral alleles to genetic variants in Lake Malawi, and built a genome browser to visualise genome wide datasets for East African cichlids.

During an initial analysis of the Lake Malawi dataset I found that heterozygosity based  $N_e$  estimates range from  $\sim 7,200$  to  $\sim 24,300$ , and that  $F_{ST}$  between Lake Malawi species varies between 0.04 and 0.66. A phylogenetic approach to the study of species relationships provides a strong signal when averaging across all genomic loci, but there is a lot of discordance between local phylogenies, and especially between phylogenies built using nuclear and mitochondrial DNA data. The discordance is due to high prevalence of incomplete lineage sorting, but interspecific introgression may also play a role. In three cases of possible introgression I formally tested the hypothesis using the ABBA-BABA statistic. In the first case, where introgression was suggested by

phenotypic similarity, I did not find any evidence. Therefore, the phenotypic similarity in denture between shallow and deep water *Lethrinops* may be a case of parallel evolution. In the other two cases, introgression was suggested by unexpectedly high co-ancestry (sharing of haplotypes). In both cases, the ABBA-BABA statistic confirmed introgression, from *P. electra* into *B. rhoadesii* (weak;  $f=1.3\%$ ) in Lake Malawi and from *O. tetrastigma* into *Astatotilapia* (strong;  $f=29.8\%$ ) in the crater lake Ilamba. While the co-ancestry relationships in the Lake Malawi dataset suggest several other possible cases of introgression between relatively closely related species, it is perhaps not as common as previously thought, and seems to be very rare between more distinct lineages - perhaps possible only under special conditions, such as in the isolation and turbid waters of Lake Ilamba.

I described a detailed analysis of early-stage adaptive divergence of two cichlid fish ecomorphs in Lake Massoko, a small (700m diameter) isolated crater lake in Tanzania. The ecomorphs differ in depth preference, male breeding colour, body shape, diet and trophic morphology. With whole genome sequences of 145 fish, I showed multiple lines of evidence that the divergence between the ecomorphs has happened in sympatry within Lake Massoko. I identified 98 clearly demarcated genomic ‘islands’ of high differentiation on the basis of  $F_{ST}$  and characterised these highly diverged regions (HDRs) in terms of local absolute sequence divergence, nucleotide diversity, and genomic location, showing that the results are consistent with the ‘islands of speciation’ model of speciation with gene-flow. I also designed a SNP assay to support aquarium mate choice experiments that demonstrate association of genotypes across these islands to divergent mate preferences. The HDRs contain candidate adaptive genes enriched for functions in sensory perception (including rhodopsin and other twilight vision associated genes), hormone signalling and morphogenesis. The possible adaptive role of rhodopsin is further supported by results of a collaboration confirming a shift in light absorption spectra between the two alleles of rhodopsin in Lake Massoko.

## 7.2 Future work

The Lake Massoko study is one of the most extensive genomic and ecological studies of divergence in the initial phases of speciation to date, especially in a sympatric setting. Next, we would like to gain a better understanding of the significance and functions of the discovered HDRs. Funding has been obtained and work is under way to collect ~200 Massoko males across several depth transects around the lake, measure anatomical and ecological traits and sequence the whole genomes for each individual fish. We will then

test for association between genetic variants and measured traits. The significance of associations will be tested in an independent sample. This work will be complemented by additional mate-choice trials under varying lighting conditions to approximate the different light environments in the lake. The ecomorphs of Lake Massoko show clear differences in traits normally associated with adaptive radiation in cichlid fishes, including body shape, pharyngeal jaw morphology, diet, microhabitat preference, retinal pigment sensitivity, male colour and mate preference [82, 83, 93, 206, 221]. Therefore, this study suggests processes and specific genomic regions that may be involved in speciation events within the great cichlid radiations of Lakes Malawi, Victoria, and Tanganyika.

After more than a century of work on the taxonomy and evolution of Lake Malawi cichlids, involving thousands of measurements of individual morphological characters, the relationships between individual species still remain unclear. Today, with whole genome sequence data, we have an unprecedented opportunity to reconstruct the evolutionary history of the radiation, as demonstrated by the initial results presented in this thesis. Understanding the causes and consequences of speciation and subsequent adaptation requires investigation of taxon pairs at different levels of divergence. Knowledge of species relationships, including the relative timing, frequency and sequence of evolution of major adaptive innovations will open up this complex and fascinating system to a new generation of research.

The first immediate steps will be adding data from the 32 samples that have been sequenced in Summer 2015 to the existing analysis. I expect this data to provide additional specific insights, in particular in relation to 1) The position of the genus *Rhampochromis* within the radiation; 2) The relative relationships of *Astatotilapia rujeva*, *A. tweddlei*, and *Serranochromis robustus* to the Lake Malawi flock and their contributions to the Malawi haplochromine radiation; 3) The origin of the ‘large predator’, and ‘molluscivore’ phenotypes in Lake Malawi; 4) The potential source of gene flow to Lake Massoko, with four individual *A. calliptera* sequenced from the Itupi stream, the closest existing water body upstream of Lake Massoko; 5) The geographical context of the origin of two very different forms of *Rhampochromis* in Lake Kingiri.

Our data now contains multiple samples from each of the major evolutionary lineages and covers 34 genera, as described in chapter 3. There are currently 53 recognised endemic genera in Lake Malawi, several of which are polyphyletic as shown by the initial analysis in this thesis. Therefore, I estimate that ~50 additional species will need to be sampled for a satisfactory overview of Lake Malawi evolutionary history down to the generic level.

As alluded to on several occasions previously in this thesis, I am especially interested in studying the genetic basis of parallel evolution. Within Lake Malawi, there are also several adaptations present in species that are not believed to be very closely related, for example the ‘thick lip’ phenotype 7.1, or the *Lethrinops*-like denture (demonstrated in this thesis).

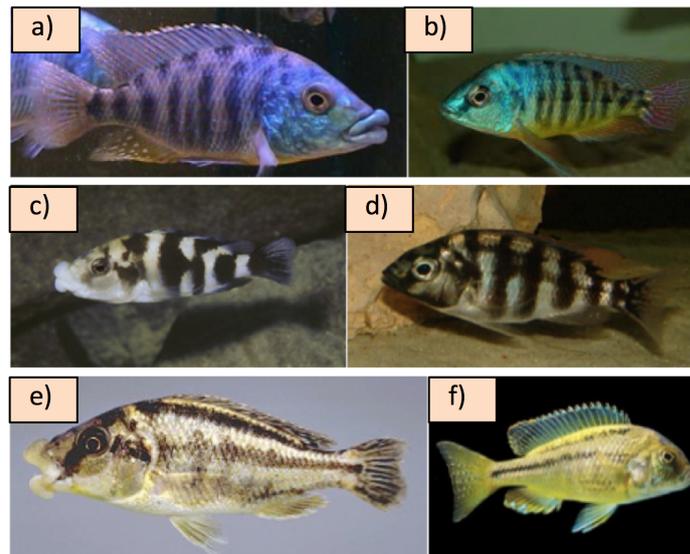


Fig. 7.1 **Parallel evolution of the ‘thick lip’ phenotype in Lake Malawi.** Contrast 1: (a) *Protomelas ornatus* (b) *Protomelas fenestratus*; Contrast 2: (c) *Placidochromis milomo* (d) *Placidochromis* ‘johnstoni solo’; Contrast 3: (e) *Chilotilapia euchilus* (f) *Chilotilapia rhoadesii*; Species a, c, and e are not believed to be particularly closely related but all possess the ‘thick lip’ phenotype. On the other hand the pairs a-b, c-d, and e-f are closely related but b,d, and f lack the ‘thick lip’ adaptation. Images from George Turner.

Do such parallel phenotypes indicate genetic parallelism? Does the adaptation involve changes in the same genes, genomic regions, or even in the same genetic variants [251]? If the same genetic variants are involved, is it because mutations that occurred independently in different species, because the allele was already present in a shared ancestral population, or because it was transferred from one species to another by introgression [252]?

In section 3.4, I describe whole genome alignments that will facilitate direct comparisons of genomic regions identified in Lake Malawi studies to genomic regions underlying speciation or adaptive divergence in lakes Tanganyika and Victoria. Thus, it will be possible to draw more general conclusions about speciation and adaptive divergence in East African cichlids, and to address the genomic basis of parallel phenotypic evolution

across lakes, which is often even more striking than within Malawi (Figures 1.12, 7.2). Walter Salzburger and his team at the University of Basel in Switzerland are currently sequencing the whole genomes of two individuals each from all of the recognised species of Lake Tanganyika with all data due to become available by summer 2016 (W. Salzburger, pers. comm.). The combination of this dataset with the Malawi dataset presented in this thesis could provide an excellent starting point for the first whole genome across-lake study in East African cichlids.

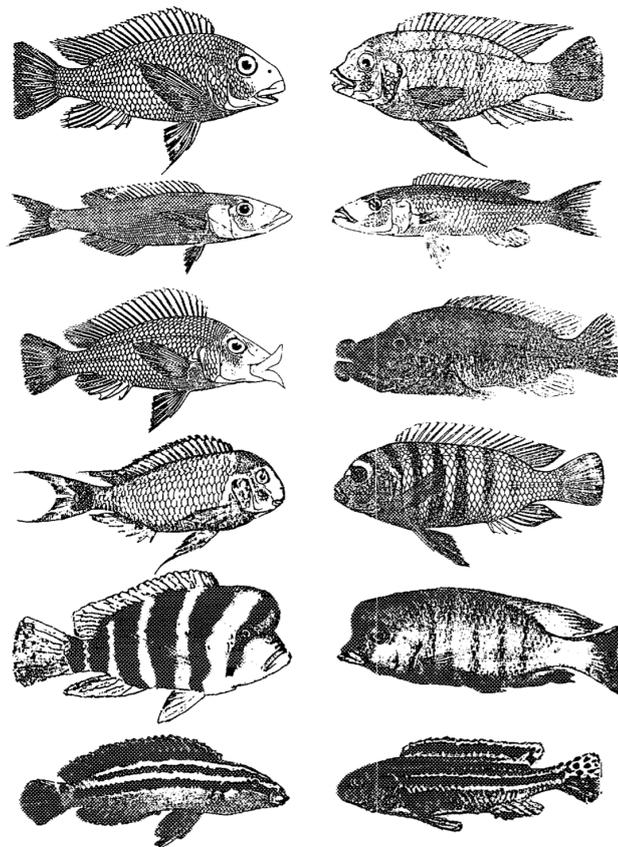


Fig. 7.2 **Parallel evolution between lakes Tanganyika and Malawi.** Six pairs of morphologically similar species from Lake Tanganyika (left) and Lake Malawi (right). The specific shared features are (from top to bottom): rasping jaw morphology, elongated body typical of pursuit predators, ‘thick lips’, mbuna habit, enlarged nuchal hump, and strong horizontal striping. Image from Kocher *et al.* [90].